DOCUMENT RESUME

ED 139 658 SE 022 533

AUTHOR Coldeway, Nancy A.; Merrill, M. David
TITLE Attribute Isolation Versus Correct Answer Feedback Effects on Response Confidence, Latency, and Test Performance. BYU Instructional Science Report No. 52.
INSTITUTION Brigham Young Univ., Provo, Utah. Coll. of Education.
REPORT NO BYU-ISR-52
PUB DATE Nov 76
NOTE 43p.; Paper presented at the annual meeting of the American Educational Research Association (New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS College Students; *Concept Formation; *Educational Research; *Feedback; Higher Education; *Instruction; *Learning; *Science Education
IDENTIFIERS Research Reports

ABSTRACT

Twenty-nine college undergraduates completed practice problems to learn the concept "Tachaic meter", receiving either no feedback (NF), correct answer feedback (CA), or attribute isolation feedback (AI). Subjects recorded their confidence in their answers for both practice and test items, along with the amount of time they spent studying the practice feedback (in CA and AI conditions). On practice items, AI feedback produced greater confidence than NF. The subjects' confidence increased after the first six (out of eighteen) practice items. More time was spent on feedback following an incorrect response than a correct response, and this difference increased as confidence increased. For a correct answer, time decreased as confidence increased and overall, more time was spent on AI feedback than CA feedback. For an incorrect answer, this direction reversed for AI and CA feedback conditions from medium to high confidence. On the test, AI and CA feedback conditions had more correct answers and higher confidence than the control NF group. Implications for the design of instruction are discussed. (Author/DT)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

structional science report instr
report instructional science rep
cience report instructional scie
nal science report instruction
nstructional science report in

ED139658

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BYU INSTRUCTIONAL SCIENCE REPORT

ATTRIBUTE ISOLATION VERSUS CORRECT ANSWER FEEDBACK EFFECTS ON RESPONSE CONFIDENCE, LATENCY, AND TEST PERFORMANCE

Nancy A. Coldeway & M. David Merrill

BYU INSTRUCTIONAL SCIENCE REPORT
NO. 52

2

BRIGHAM YOUNG UNIVERSITY • COLLEGE OF EDUCATION
Provo, Utah 84602

SE 022 533

ATTRIBUTE ISOLATION VERSUS CORRECT ANSWER FEEDBACK EFFECTS ON RESPONSE CONFIDENCE, LATENCY, AND TEST PERFORMANCE

Nancy A. Coldeway
M. David Merrill

REPORT NO. 52
November 1976

This report is a working paper designed for prepublication distribution. The reader's comments are solicited and should be conveyed directly to the author(s) or the editor. Please do not cite as a source without consulting the editor or the author(s) of this document.

BYU Instructional Science Reports have replaced the Instructional Research and Development Working Paper Series previously published by Brigham Young University Division of Instructional Services. Instructional Science Reports are distributed to faculty members and students in the Instructional Science Department and to a limited number of interested professionals in the field of Instructional Science. If you would like to receive BYU Instructional Science Reports, please notify the editor.

BYU INSTRUCTIONAL SCIENCE REPORT
Department of Instructional Science
College of Education
Brigham Young University
Provo, Utah 84602

M. David Merrill, Editor
Nan L. Howard, Production Editor

CONTENTS

Abstract 1

Method 6

Task 6

Apparatus 9

Treatments 9

Test 10

Procedure 10

Design 10

Subjects 14

Results 14

Discussion 14

Confidence on practice items as a dependent variable 19

Time spent on feedback as a dependent variable . . . 20

Additional findings re: time spent on feedback as a dependent variable 21

Performance on test as a dependent variable 23

Confidence on test as a dependent variable 23

Reference note 27

References 28

Footnotes 30

Abstract

Twenty-nine college undergraduates completed practice problems to learn the concept "Trochaic meter," receiving either no feedback (NF), correct answer feedback (CA), or attribute isolation feedback (AI). Subjects recorded their confidence in their answers for both practice and test items, and the amount of time they spent studying the practice feedback (in CA and AI conditions). On practice items, AI feedback produced greater confidence than NF. The subjects' confidence increased after the first six (out of eighteen) practice items. More time was spent on feedback following an incorrect response than a correct response, and this difference increased as confidence increased. For a correct answer, time decreased as confidence increased and overall, more time was spent on AI feedback than CA feedback. For an incorrect answer, this direction reversed for AI and CA feedback conditions from medium to high confidence. On the test, AI and CA feedback conditions had more correct answers and higher confidence than the control NF group. Implications for the design of instruction are discussed.

Attribute Isolation versus Correct Answer Feedback Effects on Response Confidence, Latency, and Test Performance

This study was designed to test the effectiveness of feedback variables in promoting concept learning. The purpose of this study was to extend the research done by Kulhavy, Yekovich, and Dyer (in press). In addition to correct answer feedback, attribute isolation feedback was used in the present study in an effort to improve concept learning. Additionally, the variable of response confidence, which was used as an independent variable by Kulhavy, et.al. (in press), was also measured as a dependent variable in this study. The elaboration of these and other variables follows.

The independent variables investigated in this study were (a) Type of Feedback, (b) Number of Practice Items Completed, (c) Response Confidence, and (d) Correctness of Response. The three types of feedback used were attribute isolation (AI), correct answer (CA), and no feedback (NF). Attribute isolation was defined as feedback which immediately followed a response, indicated whether the response was correct or not, and then isolated all of the critical attributes of the concept (Merrill & Boutwell, 1973; Merrill & Tennyson, in press; Merrill & Wood, 1974, 1975). Correct answer feedback only indicated whether the answer was correct without any further elaboration of the critical attributes of the concept, as in Kulhavy, et.al. (in press).

The no feedback treatment differed from that of Kulhavy, et.al. (in press) in that this group not only received no feedback, but they also received no definition of Trochaic meter. This was done specifically to determine if a ceiling effect would result from knowledge obtained prior to the experiment. This group served as a control for prior knowledge.

The number of practice items completed were grouped into three levels: items 1-6, items 7-12, and items 13-18. The items were analyzed by groups because separate analysis of all 18 practice items exceeded the capacity of the computer program.

Response confidence refers to the learner's self report of his/her confidence in the correctness of the previous response on a practice item (Kulhavy, et.al., in press). Practice items confidence served as an independent variable influencing the amount of time spent studying feedback.

Finally, the learner's response to an individual problem was identified as either correct or incorrect. Errors were also treated as an independent variable influencing the amount of time spent studying feedback.

Four dependent measures were recorded. These were (a) Confidence on Practice, (b) Time Spent on Feedback During Practice, (c) Performance on Test, and (d) Confidence on Test. Confidence on practice was measured as a dependent variable which was a function of the type of feedback and

number of practice items completed. The confidence variable is viewed as a dependent or independent variable, depending on the point in time. This warrants further elaboration. On two successive practice problems with feedback, confidence (dependent variable), as measured on the second problem, is a function of the type of feedback received on the first problem. It is also predicted that the dependent variable of confidence will be influenced by the number of practice items which have been completed by the learner. Confidence functions as an independent variable on every practice problem in that it (along with several other independent variables) influences the amount of time which will be spent on feedback.

The second dependent variable, time spent on feedback during practice, was a measure of how many seconds each individual spent studying each piece of feedback in the AI and CA conditions (Kulhavy, et.al., in press).

The dependent variable of performance on the test was a simple count of number of correct answers on the test (Kulhavy, et.al., in press).

The dependent measure of confidence on the test items asked students to state their confidence for each of their responses on the test, just as they had done on practice items (Kulhavy, et.al., in press). The test measured the ability to learn the concept, as opposed to memory or rule using ability.

The first extension of Kulhavy, et.al., (in press) was on the dependent variable. A concept test requiring identification of previously unencountered instances replaced the memory level test which contained the same items as used during practice. This change was made to extend the generalizability of the results to concept learning (Merrill & Tennyson, in press).

The second extension was the addition of the attribute isolation feedback condition. Attribute isolation feedback has been shown to produce better performance on tests than correct answer feedback (Tennyson, Steve, & Boutwell, 1975; Young, 1972), and it has also produced more efficiency in learning and test performance (Tennyson, et.al., 1975).

For this study it is hypothesized that:

1(a). AI feedback will produce higher confidence on practice items than CA feedback. Both of these conditions are expected to produce higher confidence than the NF condition.

1(b). Confidence will be lower for practice items 1-6, than 7-12 and 13-18; and confidence will be lower for practice items 7-12 than 13-18.

1(c). There will be an interaction between the type of feedback and the number of practice items completed. As the number of practice items completed increases,

confidence will increase more for AI than CA feedback conditions, and confidence will be lowest for the NF condition.

2(a). More time will be spent on feedback during practice after an incorrect response, than a correct response (Anderson, Kulhavy, & Andre, 1971; Kulhavy, et.al., in press).

2(b). The AI condition will spend more time on feedback than the CA condition.

2(c). There will be an interaction between confidence and correctness of response, with no difference between correct and incorrect responses for low confidence, a difference between correct responses for medium confidence, and an even greater difference for high confidence.

3. The AI feedback group will have more correct answers on the test than the CA feedback group and the CA feedback group will have more correct answers than the NF group.

4. Confidence on test items will be highest for the AI group, lower for the CA group, and lowest for the NF group.

Method

Task. The task for this study was the classification of segments of poems (ranging from one to five lines in

length) as either trochaic or not trochaic. The poems were centered on 8½ x 11-inch pieces of paper. The first page of the materials asked if the learner already knew how to scan for poetic meter. The next three pages were the directions and a page explaining the importance of the study. In the AI and CA feedback treatments, the fifth page gave directions for determining if a poem was trochaic (Figure 1). The no feedback condition did not have this page. The next page (in all conditions) gave directions for recording time data in minutes and seconds.

Insert Figure 1 about here

Then the practice poems followed. The two questions following each poem were (1) "Is this poem trochaic?" followed by a box with "+ -" over it as a reminder that the directions had stated that a "+" was to be put in the box if the poem was trochaic and a "-" was to be put in the box if the poem was not trochaic; and (2) "How sure are you?" followed by a hexagon with "1234" over it as a reminder that if you guessed the answer, mark "1," if you were somewhat unsure of the response, mark "2," if you were somewhat sure, mark "3," and if you were confident of your answer, mark "4." The upper right corner of every page had a box to mark the time when each page was started. In the AI and CA feedback

treatments each practice poem was followed by a blank piece of paper (to prevent the learner from reading through the page to the answer) and a page of feedback.

Following the eighteen practice items (half positive instances and half negative instances, randomly ordered and given to all treatments in that order), was a page stating that eighteen test poems were to follow and a repetition of the "+" and "-" directions for indicating if the poem was trochaic and the confidence scale "1234." For the two feedback conditions, the next page indicated that the directions page (Figure 1) was to be returned to the booklet at this time and not used on the test. Time directions were repeated on the next page. Eighteen randomly ordered (same order used in all conditions), previously unencountered test items followed. The last page indicated that it was the end of the program and had a place for recording the last time data.

The poems used were taken from Osguthorpe (1975), Merrill and Tennyson (in press), and Woolley (1971). Wherever data were available, an effort was made to include six easy, six medium, and six hard instances (half each of positive and negative instances) for the practice and test items. The instance probability analysis of Woolley (1971) was used for this data. The mixture of easy, medium, and hard instances was used based on the results of the study

by Merrill and Tennyson (Note 1) which showed that correct classification resulted more often under these conditions than when any other combination of difficulty of instances was used.

Apparatus. The booklets were held together by one-inch rings so the pages could be flipped over. An electronic digital clock presented the time in one-inch numbers illuminated in red. The clock displayed hour, minute and second.

Treatments. Three treatments were used to investigate the hypotheses given in the introduction of this study. The attribute isolation (AI) feedback treatment group received detailed feedback which pointed out the critical attributes. The feedback for the AI treatment had this key in the upper left corner: ___ = stressed syllable, / = division between feet, ◯ = a trochaic foot. In the poem itself, the syllables were hyphenated, the stressed syllables underlined, the feet divided by slashes, and the trochaic feet (stressed syllable followed by unstressed syllable) circled. (See Figure 2.) A statement was then made as to whether most of the feet in the poem were trochaic or were not trochaic and a "+" or "-" was in the box below the appropriate statement that "This poem is trochaic" or "This poem is not trochaic."

The correct answer (CA) treatment simply had the poem retyped on the feedback page along with the statement "This poem is trochaic" or "This poem is not trochaic." The box below had the appropriate "+" or "-" in it. The final treatment group received no feedback on the practice items. One practice poem immediately followed another.

Insert Figure 2 about here

Test. Immediately following the treatments, each student began the test. The eighteen unencountered test items were scored as correct or incorrect classifications of the poem as trochaic. Time data were recorded in the upper right corner of every page. No feedback was given during the test and no directions could be used to help answer the question "Is this poem trochaic?" The test items were written in the same format as the practice items.

Procedure. All learners received the treatment as a group in a one-hour block of time. The clock was displayed prominently in the front of the room. Three different treatment conditions were stacked in an alternating AI, CA, NF order, so that when they were distributed, there would be approximately the same number of students receiving each treatment. The students were directed to put everything away except for a pencil. The booklets were then passed

out and the following directions were read aloud by the experimenter.

"Put your name and your professor's name on the top sheet of the paper so you will receive credit for participating in the experiment. Circle your class standing. Turn the page and answer the question, "Do you know how to scan for poetic meter?" For the purposes of this research study, four things are very important: (1) do not look at anyone else's paper, (2) answer all questions as best you can, but answer everything, (3) record time accurately, and (4) do not turn back once you have turned the page. Work steadily, you will have all the time you need. Now try recording the time on this clock for practice. (Experimenter writes time on board and explains hour, minute, and second division where the time data is recorded.) When you have finished, close your booklet and sit quietly until everyone is finished. Are there any questions? Begin."

When the last student finished, the booklets were collected and the students were debriefed on the study. Because of the size of the room and the manageable number of students, the experimenter was able to watch carefully to ensure that no one looked back through the instruction during the experiment, and that direction pages were put away before taking the test.

Design. This study can be broken down into three interrelated designs. The first corresponds with hypotheses 1(a), (b), and (c). The second is based on hypotheses 2(a), (b), and (c). The third design is for hypotheses 3 and 4. All three designs follow the Campbell and Stanley (1963) posttest only control group design. The NF treatment serves as the control group.

The first design analyses the effect of type of feedback and number of practice items completed on confidence in practice items. It is a 3 x 3 repeated measures design with three levels of feedback (AI, CA and NF), three levels of practice (items 1-6, 7-12, and 13-18), and the repeated measure is across subjects per group (10 in AI feedback and NF groups, and 9 subjects in the CA feedback group). The dependent variable was the subjects' confidence on the practice items. A univariate analysis of variance was

performed. Newman-Kuels tests were used to interpret the feedback and practice effects because the model included the random factor of subjects. The Newman-Kuels tests compared AI and CA feedback versus NF, AI versus CA feedback; and practice items 1-6 versus 7-12 and 13-18, 7-12 versus 13-18.

For the second set of hypotheses (time on feedback as a function of confidence, correctness of response, and type of feedback) a 2 x 2 x 3 design using univariate analysis of variance was performed on data which had been transformed with a square root function. This transformation was necessary because the cell sizes varied from 11 to 94. There were two levels of feedback (AI and CA), two levels of correctness of response (incorrect and correct on any one item) and three levels of confidence (Low, Medium, and High, corresponding to confidence ratings of 1 and 2 combined, 3, and 4, respectively).[1] Planned comparisons were run which compared AI versus CA; incorrect versus correct; and low versus medium and high, medium versus high. The dependent measure for this hypothesis was number of seconds spent on feedback during practice.

The third hypothesis (test performance as a function of the type of feedback) and the fourth hypothesis (confidence on test items as a function of the type of

feedback) were analyzed together with the same design. A multivariate analysis of variance with three levels of feedback (AI, CA, and no feedback) and two dependent variables (number of errors on test and confidence on test) was performed. Two univariate analyses of variance were run (one on each dependent variable) and used to interpret the univariate analysis of variance results. Planned comparisons were run with the univariate tests, comparing AI and CA versus NF, and AI versus CA feedback. All analyses accounted for any missing cells due to the unequal N's in the treatment conditions.

Subjects. Twenty-nine college students from introductory psychology courses at Brigham Young University participated in the experiment. There were ten students each in the AI feedback and NF treatments, and nine students in the CA treatment. The uneven group size was due to the fact that one subject failed to appear for the experiment. All subjects completed all of the instruction and the test in less than one hour.

Results

Hypothesis 1(a) stated that AI and CA feedback would produce higher confidence than NF on practice items, and that AI feedback would produce higher confidence than CA

feedback. Figure 1 illustrates the means for hypothesis 1. This hypothesis was completely supported. The analysis of variance showed feedback to be a significant effect, $F(2,26) = 17.864$, $p < .0005$. Newman-Kuels tests showed that AI and CA feedback produce higher confidence than NF, and that AI feedback produces higher confidence than CA ($\alpha = .01$). The main effect means are: 3.433, 2.821, and 1.611 for AI, CA and NF treatments, respectively.

Insert Figure 3 about here

Hypothesis 1(b) stated that as the number of practice items completed increased, confidence would increase. This hypothesis was also supported (Figure 3). It was tested by the planned orthogonal comparisons of 1-6 versus 7-12 and 13-18, and 7-12 versus 13-18. Practice was a significant effect, $F(2,52) = 3.377$, $p < .05$. The planned comparisons show that confidence was higher on the last 12 practice items, than on the first 6, $F(1,52) = 6.53$, $p < .025$. There was no difference between practice items 7-12 and 13-18 $F(1,52) = .224$, $p > .05$. The respective means were 2.500, 2.702, and 2.663.

Hypothesis 1(c) stated that there would be an interaction between type of feedback and number of practice items completed. This was not supported by the data, $F(4,52) = .354$, $p > .05$ (Figure 3).

Hypothesis 2(a) stated that more time would be spent studying feedback after an incorrect response, than after a correct response. Figure 4 illustrates the means for hypothesis 2. This hypothesis was supported in the direction predicted, $F(1,330) = 45.05$, $p < .0005$. The means for incorrect and correct responses were, 4.212 and 3.397, respectively.

Insert Figure 4 about here

Hypothesis 2(b) stated that the AI group would spend more time on feedback than the CA group. The data also support this hypothesis, $F(1,330) = 10.266$, $p < .005$ (Figure 4). The means were: AI = 3.999 and CA = 3.610.

Hypothesis 2(c) stated that there would be an interaction between confidence on practice items and correctness of response. This was supported, $F(2,330) = 8.536$, $p < .0005$. The orthogonal planned comparisons showed one significant effect, $F(1,330) = 16.867$, $p < .0005$, i.e., as confidence increases, regardless of the type of feedback, more time is spent on incorrect responses than on correct responses (Figure 4). The amount of time spent on feedback is not significantly different for correct and incorrect responses at low confidence, $F(1,330) = .285$, $p > .05$; but the time is significantly different between correct and incorrect responses on medium and high confidence scores. The means are given in Table 1.

Insert Table 1 about here

Two additional interactions were significant. These were not foreseen, but are reported here because the analysis showed them to be significant. Both relate to the second set of hypotheses which examined the effects of type of feedback, correctness of response, and confidence on time spent on feedback.

The first significant interaction was between type of feedback and confidence, $F(2,330) = 10.611$, $p < .005$. The two planned orthogonal comparisons were significant. The first planned comparison indicates that at low confidence the AI feedback group spends more time on feedback than the CA feedback group, and this time is significantly lower than for AI and CA feedback at medium and high confidence, $F(1,330) = 7.228$, $p < .001$. The second planned comparison indicates that at medium confidence, the AI feedback group spends more time on feedback than the CA group, and at high confidence, this is reversed, $F(1,330) = 14.395$, $p < .0005$. The means are given in Table 2.

Insert Table 2 about here

The second significant interaction is the three-way interaction of type of feedback, correctness of response, and confidence, as they effect time spent on feedback, $F(2,330) = 4.217$, $p < .025$. Only the second planned comparison was significant, indicating that there was a

significant difference in the interaction of type of feedback and correctness of response between medium and high confidence, $F(1,330) = 8.220$, $p < .0005$. but that this interaction was not significantly different between low versus medium and high confidence, $F(1,330) = .270$, $p > .05$. See Figure 4 for the graph of the interaction.

Hypothesis 3 predicted that AI and CA feedback would produce better test performance than NF; and that AI feedback would produce better test performance than CA feedback. A multivariate analysis of variance was done with test performance and confidence on test (hypothesis 4) as dependent variables and type of feedback as the independent variable. Univariate analyses were run on the two dependent variables to interpret any significant multivariate results. The first part of this hypothesis was supported. The univariate test on performance showed feedback to be a significant variable, $F(2,26) = 13.317$, $p < .0005$. The first comparison (AI and CA > NF) was significant, $F(1,26) = 25.167$, $p < .0005$. The second comparison (AI versus CA) was not significant, $F(1,26) = 1.114$, $p > .05$. The respective means for AI, CA and NF are 14.700, 13.556, and 9.500.

Hypothesis 4 stated that confidence on test items would be higher for the AI and CA feedback groups than for the NF group; and that confidence would be higher for the AI than the CA feedback group. The multivariate test

reported above justifies a univariate test. Feedback was a significant effect on confidence $F(2,26) = 7.551$, $p < .005$, and the first planned comparison was significant, $F(1,26) = 12.273$, $p < .05$, i.e., AI and CA feedback produce higher confidence than NF (Figure 7). The second planned comparison (AI > CA) was not significant, $F(1,26) = 2.520$, $p > .05$. The respective means for AI, CA, and NF are 3.390, 2.890 and 2.190.

Discussion

The discussion of this research study is divided into three sections relating to the three interrelated designs.

Confidence on practice items as a dependent variable. The results of the present experiment support two conclusions regarding confidence as a dependent variable. First, attribute isolation feedback produces higher confidence than either CA feedback or NF; and CA feedback produces higher confidence than NF. Second, the learner's confidence increases significantly after the first third of practice. The support of the first two hypotheses measuring confidence on practice items indicates that learner confidence in instruction can be influenced by the type of feedback which is given. These results suggest that AI feedback be used to produce high learner confidence in instruction. The increase in confidence as practice continued is attributed to familiarity with the materials, since the increase in

confidence was a constant across all treatments (the type of feedback x practice items completed interaction was not significant). A feedback x practice interaction was initially predicted, based on the idea that confidence would increase significantly more for the AI treatment than for the CA or NF treatments. The data show that this is not the case.

Time spent on feedback as a dependent variable. The results of the second set of hypotheses (time spent on feedback) completely support and replicate the findings of Kulhavy, et.al. (in press). First, the results showing that more time was spent on feedback following an incorrect response than a correct response provide an extension of the Kulhavy, et.al. (in press) research to a concept level task. Additionally, it was found that more time was spent on feedback when students were given AI feedback than when they were given CA feedback. This is attributed to the fact that there is more information available to the learner in the AI condition. The third, and most interesting finding, also supports the Kulhavy, et.al. (in press) results, and follows their predictions that if the instruction is not understandable, the learner will have low confidence, and feedback will be of little help, regardless of the correctness of the response. For the learner who is confident that the question has been answered correctly,

there is no surprise when, in fact, the feedback confirms this expectation. In this case, feedback serves to confirm an expectation and a quick glance at the feedback is all that is needed. The time when feedback seems to be most valuable is when the learner is confident that the question has been answered correctly, when it actually has been answered incorrectly. This learner is faced with feedback that does not match the expectation. When the expectation has not been confirmed, the learner uses the feedback to analyze why the error occurred. These results indicate that feedback should be written in such a way as to anticipate common errors that might be made, and include an explanation of these errors for the confident, but erring student, since this is where feedback plays the greatest corrective role.

Additional findings re: time spent on feedback as a dependent variable. The first of the two additional significant findings was the two-way interaction between type of feedback and confidence as measured by time spent on feedback. At low confidence, the AI feedback group spends more time studying feedback than the CA feedback group and this time is less than at medium or high confidence ratings. The fact that the AI feedback group spends more time studying the feedback than the CA group

follows the main effect of AI > CA. It also indicates that less time is spent on feedback by the unconfident student. This is logical since the student with low confidence probably does not understand the instruction and does not spend much time looking at the feedback. The unusual finding here is the significance of the second planned comparison. There is a reverse between medium and high confidence. At medium confidence, the AI feedback group spends more time studying feedback than the CA feedback group, and at high confidence these groups reverse. It is not clear why, at high confidence, the AI feedback group should spend less time on feedback than the CA feedback group. Looking at the three way interaction of confidence, type of feedback, and correctness of response may help to clarify what is actually happening here.

Figure 4 shows that this flip-flop of AI and CA between medium and high confidence occurs only in the case when the learner is incorrect. When the learner is correct the graph is as predicted, i.e., time decreases as confidence increases but it decreases more for the AI feedback condition which, overall, spends more time on feedback than CA feedback condition. The unexplainable point seems to be why the AI feedback group did not continue its increasing trend across levels of confidence when incorrect. Since learners spend increasingly more time

on feedback after an error, and as confidence increases, the predicted location of the AI-high confidence point when incorrect would be somewhere above the CA-high confidence point when incorrect.

This implies that there is something about a highly confident student, who makes an error, and has been receiving AI feedback, that allows the utilization of feedback to occur more quickly than with students who have lower confidence, or who have been receiving CA feedback. Replication of this study will determine if it is a stable result.

Performance on test as a dependent varible. The results from hypothesis 3 indicate that either type of feedback (AI or CA) improves test performance over the control condition. This is not an unexpected finding, but it is not entirely as predicted since the AI feedback group did not perform better than the CA feedback group.

Confidence on test as a dependent variable. The analysis of the data on hypothesis 4 indicates that both AI and CA feedback produce higher test confidence than the control condition. This was not entirely as predicted since the AI feedback group did not have higher confidence than the CA feedback group.

AI feedback did not produce significantly better test performance or test confidence than CA feedback. When related back to the previous hypotheses it can be seen that AI feedback increases learner confidence and is studied longer by learners, but the actual test performance and confidence are the same for AI and CA feedback treatment conditions.

In the planned replication of this study, some additions and changes will be made. First, the NF condition will be divided into two conditions: one will receive instructions and no feedback, and the other will receive no instructions and no feedback. The present study included only the latter condition and the inclusion of this new group will allow for a comparison of the control group with the group which receives only instructions. Second, a knowledge of results (KOR) (right/wrong) feedback condition could be added to examine the effectiveness of this type of feedback. This was not possible in this study because the nature of the task meant that any KOR feedback was also CA feedback. Third, replication of this study would determine if any of the data were misinterpreted due to the three interrelated designs of this study. In replication, each of the three major hypotheses would be tested with separate groups of subjects. Fourth, it would be interesting to look at the increase in confidence across practice items with more

subdivisions than 1-6, 7-12, and 13-18, perhaps 4 to 6 divisions. Finally, the confidence scale used in this study was a 1 to 4 scale, and Kulhavy, et.al. (in press) used a 1 to 5 scale. A replication of this study might include the 1 to 5 scale. Our conclusions are that this study has provided strong support for the Kulhavy, et.al. (in press) results and adds new information by the inclusion of the AI feedback treatment and the concept level task.

The following is a summary of the conclusions of this study:

1. AI is greater than CA and CA is greater than NF, for confidence on practice items.

2. Confidence is higher on the last 12 practice items than on the first 6.

3. More time is spent on feedback following an incorrect response than a correct response.

4. More time is spent on AI feedback than on CA feedback.

5. Increasingly more time is spent on an incorrect answer than a correct answer as confidence increases.

6. For a correct answer, as confidence increases, time spent on feedback decreases more for the AI feedback condition which, overall, spends more time on feedback than the CA feedback condition.

7. For an incorrect answer, there is a flip-flop in the direction of AI and CA feedback conditions from medium to high confidence.

8. AI and CA feedback resulted in higher test performance than the control NF group.

9. AI and CA feedback resulted in higher test confidence than the control NF group.

Reference Note

1. Merrill, M.D. & Tennyson, R.D. Concept acquisition and specified errors as a function of relationships between positive and negative instances. Working Paper No. 29. Division of Communication Service. Brigham Young University, Provo, Utah, October 1971.

References

Campbell, D.T. & Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Company, 1963.

Kulhavy, R.W., Yekovich, F.R., & Dyer, J.W. Feedback and response confidence. Journal of Educational Psychology, in press.

Merrill, M.D. & Boutwell, R.C. Instructional development: Methodology and research. In F.N. Kerlinger (Ed.), Review of Research in Education (Vol. 1). Ithaca, N.Y.: Peacock, 1973.

Merrill, M.D. & Tennyson, R.D. Teaching concepts: An instructional design guide. Englewood Cliffs, N.J.: Educational Technology Publications, Inc., in press.

Merrill, M.D. & Wood, N.D. Instructional strategies: A preliminary taxonomy. Columbus, Ohio: Ohio State University, 1974. (ERIC Document Reproduction Service No. SE 018-771.) Revised version, Provo, Utah: Courseware, Inc., Technical Report 1R.

Osguthorpe, R.T. The effects of definition and instance presentation on the acquisition of an ordered classification rule. Unpublished doctoral dissertation, Brigham Young University, August 1975.

Tennyson, R.D., Steve, M.W., & Boutwell, R.C. Instance sequence and analysis of instance attribute representation in concept acquisition. Journal of Educational Psychology, 1975, 67, 821-827.

Woolley, F.R. Effects of the presence of concept definition, pretraining, concept exemplars, and feedback on the instruction of infinite conjunctive concepts. Unpublished doctoral dissertation, Brigham Young University, May 1971.

Young, J.I. The effects of review techniques and instance presentation on concept learning tasks. Unpublished doctoral dissertation, Brigham Young University, May 1972.

Footnotes

[1]Low and medium-low confidence ratings (1 and 2) were collapsed into one group because of one cell which only had one observation in it. This avoided the confounding of degrees of freedom with the cell with zero degrees of freedom that would have resulted.

Table 1

Means for the Interaction Effect of Confidence on Practice Items and Correctness of Response on Time Spent on Feedback

	Response	
Confidence	Incorrect	Correct
High	3.795	3.698
Medium	4.492	3.396
Low	4.348	3.097

Table 2

Means for the Interaction Effect of Type of Feedback and Confidence on Time Spent on Feedback

	Confidence		
	Low	Medium	High
AI	4.176	4.298	3.523
CA	3.318	3.590	3.922

DIRECTIONS

WHAT IS POETIC METER?

Every poem has a certain meter. It is much like the rhythm in music. Just as some notes are stressed in music, some syllables are stressed in poetry. These different stress patterns determine a poem's poetic meter. There are four types of poetic meter. In this experiment, you will only be required to learn the rhythm pattern of these types: TROCHAIC.

WHAT IS TROCHAIC METER?

When a poem is made up of a repeated pattern of a STRESSED SYLLABLE followed by an UNSTRESSED SYLLABLE, we call the poem TROCHAIC.

WHAT IS A STRESSED SYLLABLE?

A stressed syllable is that part of a word that is said LONGER, LOUDER, OR HIGHER, in pitch. Words that have only one syllable can often be stressed or unstressed depending on their position in the poetic line.

HOW DOES TROCHAIC METER COMPARE WITH THE OTHER RHYTHM PATTERNS?

The four patterns are listed below. A single repetition of a certain pattern is called a "FOOT." The different types of FEET are listed below. You will encounter all four types in this experiment, but you only need to learn the TROCHAIC pattern.

1. TROCHAIC	STRESSED SYLLABLE, UNSTRESSED SYLLABLE
2. NOT TROCHAIC	unstressed syllable, stressed syllable
3. NOT TROCHAIC	stressed syllable, unstressed syllable, unstressed syllable
4. NOT TROCHAIC	unstressed syllable, unstressed syllable, stressed syllable

Figure 1. Directions for AI and CA treatments for determining if a poem is trochaic.

START TIME [:]

Key

_____ = Stressed syllable

/ = Division between feet

( ) = A trochaic foot

(pan - sies) / (li - lies) / (king - cups) / (dai - sies) /

EVERY FOOT IN THIS POEM IS TROCHAIC

THIS POEM IS TROCHAIC

+ -

[+]

Figure 2. An example of attribute isolation feedback.

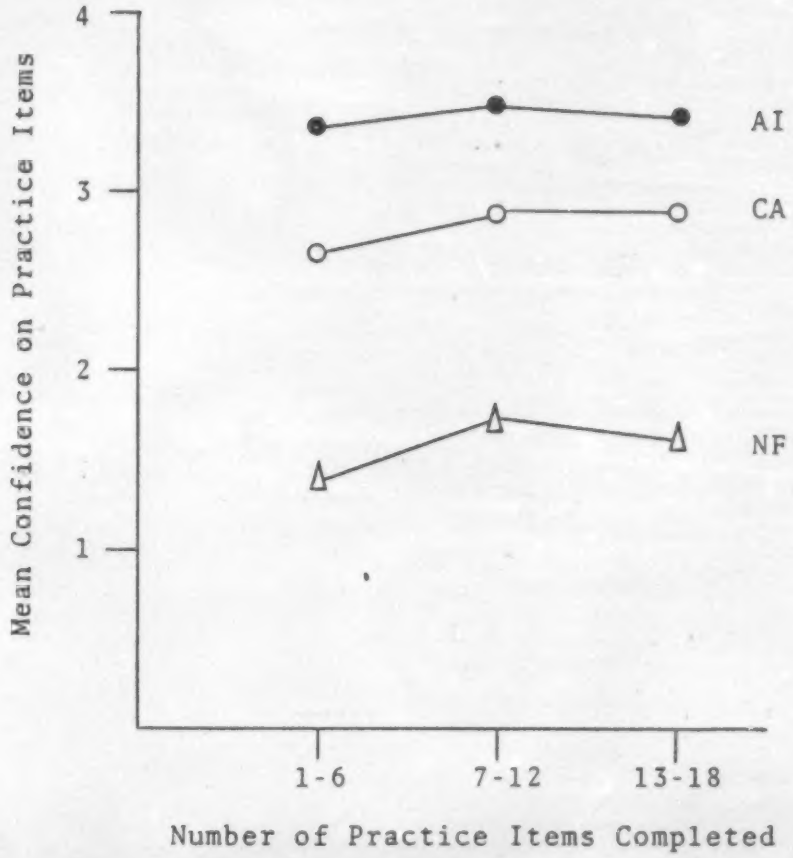

Figure 3. The effect of type of feedback and number of practice items completed on practice item confidence.

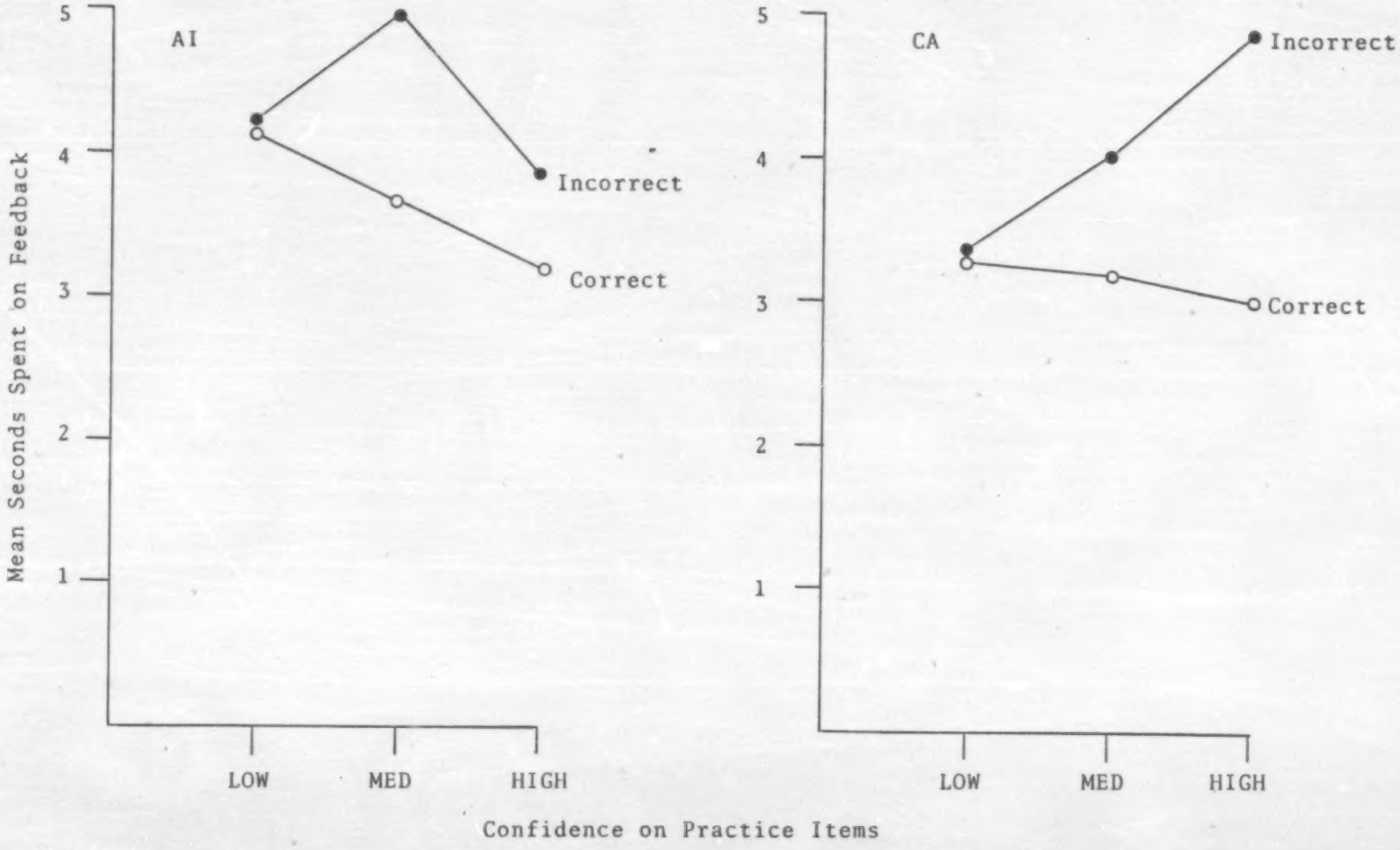

Figure 4. The effect of correctness of response, type of feedback, and confidence on time spent on feedback.